Bioquality hotspots in the tropical African flora

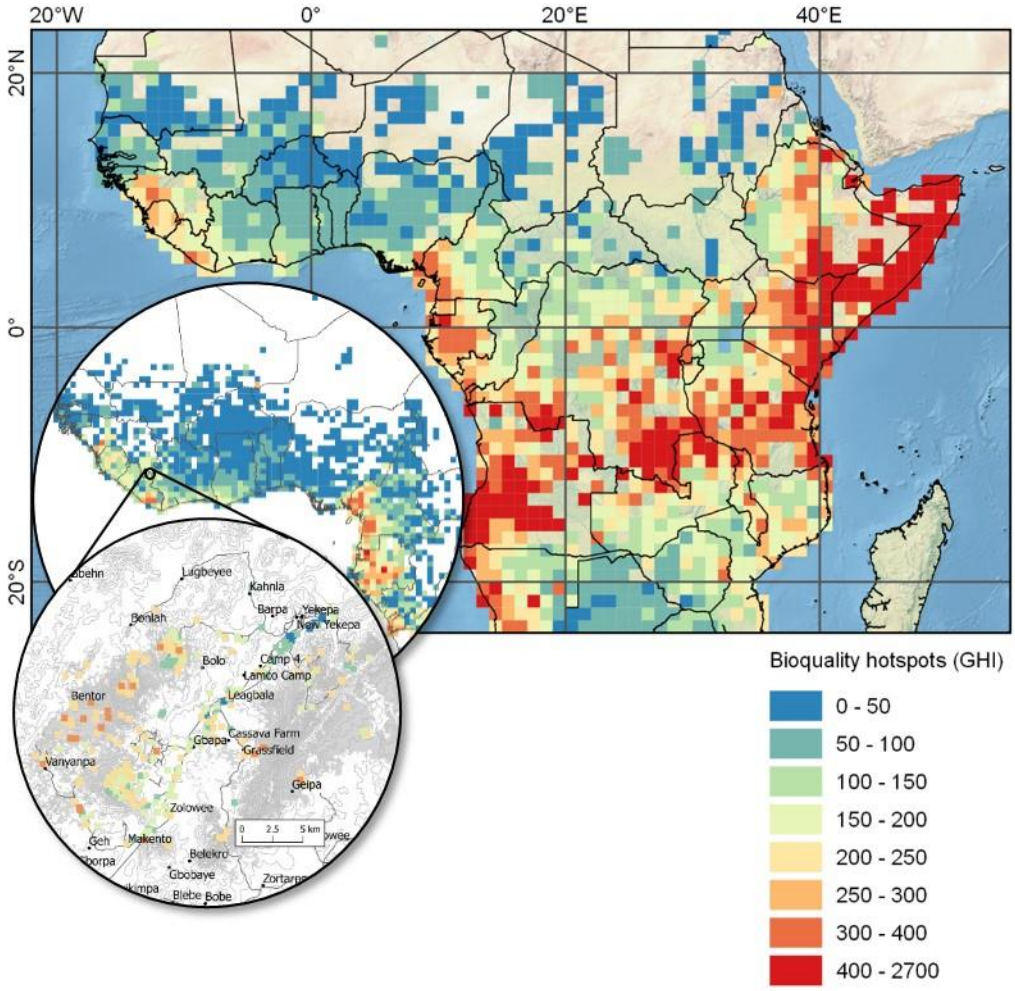
Cicely A. M. Marshall[1*], Jan J. Wieringa[2] & William D. Hawthorne[1]

[1]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom.
[2]Naturalis Biodiversity Center, National Herbarium of the Netherlands, Darwinweg 2, 2333 CR Leiden, The Netherlands.
*Correspondence to: Cicely Marshall, email: cicely.marshall@plants.ox.ac.uk; Telephone: 01865 275 140 or 01865 244 460.

Summary

Identifying areas of high biodiversity is an established way to prioritize areas for conservation [1–3], but global approaches have been criticized for failing to render global biodiversity value at a suitable scale for local management [4–6]. We assembled 3.1 million species distribution records for 40,583 vascular plant species of tropical Africa from sources including plot data, herbarium databases, checklists, and GBIF, and cleaned the records for geographic accuracy and taxonomic consistency. We summarised the global ranges of tropical African plant species into four, weighted, categories of global rarity called Stars. We applied the Star weights to summaries of species distribution data at fine resolutions to map the bioquality (range restricted global endemism) of areas [7]. We generated confidence intervals around bioquality scores to account for the remaining uncertainty in the species inventory. We confirm the broad significance of the Horn of Africa, Guinean forests, coastal forests of east Africa, and Afromontane regions for plant biodiversity, but reveal also the variation in bioquality within these broad regions and others, particularly at local scales. Our framework offers practitioners a quantitative, scalable and replicable approach for measuring the irreplaceability of particular local areas for global biodiversity conservation, and comparing those areas within their global and regional context.

Graphical abstract



Highlights

- All plant species in the region were Star rated (categories of global rarity)
- A species distribution database was assembled for tropical African plants
- A reliable minimum estimate of global irreplaceability was mapped across the region
- The results allow global conservation values to be translated into local action

In Brief

Marshall et al. introduce a new conservation framework for tropical Africa. The authors use ‘‘big data’’ to integrate species-level conservation assessments into reliable minimum local estimates of global irreplaceability across the region, providing a framework for conservationists and researchers applicable at the local scale.

Results & Discussion

Distribution data for tropical African plants

Biodiversity hotspots were originally identified using the richness of species endemic to large, biogeographic realms which had been significantly degraded [1], largely because species distribution data were available only at this coarse resolution [8]. This situation has improved rapidly as online public repositories (e.g. GBIF), collection digitisation efforts (e.g. JStor's Global Plants Initiative), and data journals (e.g. Check List) were established, increasing the available number of geolocated species records.

We assembled 3.1 million global species distribution records for tropical African vascular plants, from plot data, herbarium databases, checklists, and the Global Biodiversity Information Facility (GBIF). We limited our GBIF search to records supported by herbarium specimens and those without reported geographic issues. Our tropical African species list was derived from the African Plants Database, and includes 40,583 accepted species or intraspecific names which were checked for synonymy and comprehensiveness against other resources. We refer to species for simplicity, but all analysis was conducted on the lowest named taxonomic unit at or below the species level.

Of the 3.1 million distribution records, 0.5 million specimens were collected without coordinates. We geolocated these records by comparing the text locality information provided in the collectors' notes to standardised gazetteer dictionary files, and assigned to the records either point-with-radius, or polygon, coordinates, depending on the detail available in the notes. Records assigned polygons were included in the following analyses if they fitted inside the sampling units in question. We used similar geolocation methods to detect records for which the supplied coordinates and supplied text locality information conflicted; such records were checked and corrected by hand or were omitted from analysis. Many older specimens, often including types, were collected without coordinates. Using these novel methods we were able to compile records for almost all vascular plant taxa present in tropical Africa, ensure taxonomic consistency and geographic accuracy, and respect the geographic resolution of the original collection in the analysis.

We estimated the completeness of our species distribution data by comparing our species sampling levels against published estimates of species richness [9] (Figure 1A). There are many areas for which species sampling is far from complete, particularly for central Africa [10]. We must continue our efforts to fill these data gaps, but we cannot afford to ignore the biogeographic signal present in existing data or the plants we seek to record will be gone.

Star rating: Species-level conservation assessment

We summarised the global range for all plant species in tropical Africa into 4 categories of global range, called Stars [7] (Figure 2). Globally rare species are the important elements of biodiversity to conserve locally, in order to conserve species richness globally. Black Star species have the narrowest global ranges (c. 2.7 degree squares occupancy on average), Green Star species are the globally commonest (c. 72 degree squares), and Gold and Blue Star species are intermediate. Star ratings are species-specific, mutually exclusive and globally applicable, so that each species or intraspecific taxon in the world can have only one Star. Global ranges were categorized, rather than using a continuous occupancy metric, to produce a memorable framework which retains the necessary subtlety to reveal robust biogeographic patterns. Given that the full degree square occupancy of all species globally is not yet known (Figure 1A), the categorical system also allows for interpretation of the appropriate Star rating for species which are inadequately represented in herbaria, for example due to geographic or ecological biases in collections. We reviewed each species' Star in light of the best available information from online floras and other botanic resources, unless it was already a Green Star species (globally widespread). Although this introduces a degree

of subjectivity to the system, the results better reflect the true breadth of knowledge regarding species' distributions than a strict reliance on digitised records would. Each Star category carries a weight which is inverse to the mean range (measured as degree square occupancy) for all the included species of that Star category, so that rarer species and Stars have a higher weight (see Supplemental Experimental Procedures).

Star rating can be compared with the IUCN Red Listing approach when criterion B2 (AOO) is invoked [11], but Star rating requires no explicit measure of population change, regional ratings are not necessary or allowed, and the grid size for AOO calculations is standardised to one degree square (or 100 x 100 km, whichever is larger), for all plant species. Globally, three times as many vascular plant species have a Star rating compared with a Red List Category (62,868 cf. 20,147; 100% cf. 8% tropical African plant species assessed). Star rating offers a biologically pure assessment of a species' range which is relatively fast to conduct, and is useful for scientific analyses of distribution patterns as well as conservation assessment. As a consequence of this study, all tropical African vascular plant species have a Star rating, so the system can now be used to support or prioritise conservation anywhere in tropical Africa, and could be extended to other taxa.

Bioquality hotspots in tropical Africa

We used the Star ratings and species distribution summary tables to produce a quantitative measure of plant biodiversity value for areas across tropical Africa. Indexes respecting species global ranges reflect a particular component of what specialists tend to recognise as the biodiversity value of a place. We refer to this attribute of plant biodiversity as *bioquality*, and the particular index used to measure bioquality is the Genetic Heat Index (GHI) [7,12]. GHI is calculated for a unique species list for an area, by averaging over the weights of the Stars for those species found in the area. An area with a high proportion of globally rarer species in its flora achieves a high GHI and a high bioquality hotspot score.

This is similar to calculating range-size rarity [13,14], except we measured ranges globally rather than within the study area, to produce scores which are comparable globally. Range size rarity uses the continuous degree square occupancy of species, whereas we have binned ranges into the four Star categories to produce results which are not artificially precise, given that the full degree square occupancy of all species is not yet known (Figure 1A). The biggest difference is that the GHI divides by the number of species present (to produce a weighted average), which means that the GHI does not measure richness or diversity. This has the possible disadvantage that areas with high absolute numbers of rare species achieve lower GHI scores if they also include many common species, but a number of significant advantages: Areas are not downgraded if their species inventory is not complete, making the measure robust to missing data. GHI scores decrease where vegetation is invaded by globalized species. Species richness increases with the size of area under consideration: ignoring richness means that GHI scores can be calculated and meaningfully compared for areas of any shape or size, including the very local.

To conserve species globally, it is not important to prioritise individual areas with high species richness. Rather, it is important to protect areas where a high proportion of the individuals belong to globally rare species, otherwise those species would be lost from the global species pool [15]. In fact, the number of species in an area, whether rare, threatened or simply present (richness), is now generally recognised as a poor metric for identifying conservation priorities, because richness alone reveals little more than the availability of data, the size and shape of the area under consideration [5], and the biome type.

When GHI is calculated from an essentially complete species list for an area, then confidence intervals are not necessary. Neither would they be necessary if species were sampled incompletely

but representatively with respect to the true balance of Stars in the full flora, because as a weighted average the GHI includes no measure of richness. However, we cannot tell whether the recorded flora is currently biased towards the globally rarest (or commonest) species. We therefore estimated bootstrapped confidence intervals for the GHI for each degree square, given the apparent GHI (Figure 1C) and current estimated species sampling completeness (Figure 1A), to produce a confidence interval within which the true GHI value of each area is expected fall, even if sampling is currently biased with respect to Star (Figure 1B and 1D). This is one way in which uncertainty can be quantified and reliable conclusions drawn, whilst the species inventory is incomplete.

Figure 1B reveals tropical Africa's biodiversity patterns in their most complete, repeatable, and intimate detail yet. On the whole, the results fit comfortably with previous studies of the distribution of Africa's plant biodiversity [13,16–20], by highlighting the generally rather low endemism in the Sahara, Sahel and Sudanian regions, and medium to high endemism for the Guineo-Congolian, Zambezian, Somalia-Masai, Karoo-Namib, Zanzibar-Inhambane and Afromontane regions. The Somalia-Masai (Horn of Africa) flora comes out as one of the hottest floras in tropical Africa; while the large number of endemic species has been recognised [21], Somalia's high bioquality has perhaps been underappreciated relative to Africa's wetter and montane forest regions [13], most likely due to undersampling and relatively lower species richness.

Smaller scale bioquality hotspots are visible around Mount Cameroon, Mount Mulanje and Mount Chimanimani. In Guineo-Congolia, bioquality peaks in the high rainfall forests of Cameroon and Gabon towards the coasts, is higher for western Upper Guinea than in the east, and bioquality is somewhat lower but comparable for Congolia (though data are sparser). Bioquality peaks in the Zambezian region in south eastern Democratic Republic of Congo, and in southwest central Angola. For the Karoo-Namib, the coastline of southern Angola is particularly hot; the flora of the eastern coast of Africa (Zanzibar-Inhambane regional mosaic) is particularly hot in south east Tanzania.

Bioquality at local scales

Our bioquality metric (GHI) is based on a weighted average of globally rare plants, and as proportions scale meaningfully with richness and area, the scale (grain) and shape of sampling units for an analysis can be matched to its application. The data for such fine-scale bioquality analyses can be derived for a project area by on-the-ground sampling. In particular, Rapid Botanic Survey is a botanical survey technique specifically designed to collect this information with the minimum possible effort (see Supplemental Experimental Procedures), although a meaningful GHI score can be calculated from any reasonably taxonomically complete survey data e.g. relevés or all-species transects [22].

Figure 3 reveals the local variation in bioquality found by local sampling within one of these degree squares, around Yepeka (Nimba mountains, northern Liberia), and across different vegetation types and altitudes. Such local-scale information is particularly useful for land management planning. Bioquality around the Nimba mountains is lower for the more populated, lowland area around the central road corridor, and peaks in the closed canopy slope forests at higher elevations, with some variation apparent even within this forest type. It is clear that this 'hotspot' at the one degree square scale is a patchwork of hot and cold spots at a finer scale. It is useful to be able to measure how hot an area is at this rather local scale, because it is at this scale where decisions impacting biodiversity are often taken.

The background map shows minGHI at 0.5 x 0.5 degree square resolution, and reveals bioquality patterns in greater detail than the one degree map of Figure 1B, although fewer data points can be resolved to this higher resolution grid.

Bioquality as a conservation framework

Bioquality is measured using the global range of plant species. Vascular plants are often used as an indicator taxon for biodiversity measurements because they are relatively well known taxonomically and geographically, and define the terrestrial habitats in which other taxa live. If high bioquality is used to define priorities for conservation, or to inform local land management, it makes sense to consider many other aspects of an area [23], including species other than plants [24], social factors [25], economic cost/benefit analyses [26], ecosystem-wide benefits [27], phylogenetic diversity and evolutionary processes [28], and rates or risk of habitat loss [1]. We keep such measures out of our plant bioquality analysis, and promote viewing them as independent GIS layers, because mixing criteria in a single metric makes results harder to interpret and to make globally consistent. We accept that the proportion of globally rare plant species in a flora is by no means the only important factor when designing a land management plan, but it is a critical one.

As a consequence of this study, all mainland tropical African plant taxa have a Star rating and GHIs can now be calculated easily anywhere in tropical Africa where the species composition is at least partly known. This should prove useful in the context of Environmental Impact Assessment, or Protected Area planning, because a local scale hotspot map and database can: Describe a baseline; inform the positioning of infrastructure or protected areas; identify appropriate offset areas; allow precise monitoring of impacts and changes through time (with resurvey); and help devise management plans for the globally rarest species. We accept as a premise of the system that the data are never complete, and that taxonomic boundaries also shift, so the system is built to be robust in light of new information.

As much as 79% of Earth's land surface has now been prioritized for conservation under one system or another [8], and we do not wish to define yet another set of broad areas of conservation importance. Instead, our framework offers conservationists and land managers a quantitative and replicable approach for measuring the irreplaceability of particular local areas for global biodiversity conservation, and comparing those areas within their global and regional context.

Author contributions

Conceptualisation: C.M. and W.H.; Methodology and Software: C.M and W.H.; Resources and Data Curation: C.M. J.W, W.H.; Writing – Original Draft: C.M.; Writing – Review & Editing: C.M, J.W., W.H.

Acknowledgments

Cyrille Chatelain provided the African Plants Database Tropical Africa species list. Data capture, editing and broadcast of the African Plants Database is the product of a collaboration between the South African National Biodiversity Institute, the Conservatoire et Jardin botaniques de la Ville de Genève, Tela Botanica and the Missouri Botanical Garden. BISAP (Biogeographical Information System on African Plant Diversity) has been established at the Nees Institute, University of Bonn in the context of BIOTA Africa project together with several external partners, especially Jon Lovett (University of Leeds) and Peter Linder (Zurich). For more details about the dataset and a full acknowledgement of all contributors please refer to Linder et al. 2005 [29] and Küper et al. 2006 [30]. We thank M. Swaine for data and discussions; S. Harris, A. Hector, L. Turnbull, L. Hill, Z. Goodwin and R. Scotland for comments and discussions; D. Filer and A. Liddell for the TOPO website and the Star Server. Survey work in Nimba County was funded by ArcelorMittal Liberia and Euronimba Liberia Ltd in the context of Environmental and Social Impact Assessments. C.M. acknowledges support from the Clarendon Fund and Merton College; W.H. acknowledges support from the James Martin 21st Century School and BP Biofuels.

References

[1] Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature *403*, 853–858.

[2] Olson, D.M., and Dinerstein, E. (2002). The Global 200: Priority Ecoregions for Global Conservation. Ann. Mo. Bot. Gard. *89*, 199.

[3] Stattersfield, A. J., Crosby, M. J., Long, A. J. and Wege, D. C. (1988). Endemic Bird Areas of the World: Priorities for Biodiversity Conservation. BirdLife International (Cambridge, UK).

[4] Mace, G.M., Balmford, A., Boitani, L., Cowlishaw, G., Dobson, A.P., Faith, D.P., Gaston, K.J., Humphries, C.J., Vane-Wright, R.I., Williams, P.H., et al. (2000). It's time to work together and stop duplicating conservation efforts …. Nature *405*, 393–393.

[5] Brummitt, N., and Lughadha, E.N. (2003). Biodiversity: Where's Hot and Where's Not. Conserv. Biol. *17*, 1442–1448.

[6] da Fonseca, G.A.B., Balmford, A., Bibby, C., Boitani, L., Corsi, F., Brooks, T., Gascon, C., Olivieri, S., Mittermeier, R.A., Burgess, N., et al. (2000). … following Africa's lead in setting priorities. Nature *405*, 393–394.

[7] Hawthorne, W.D. (1996). Holes and the sums of parts in Ghanaian forest: regeneration, scale and sustainable use. Proc. R. Soc. Edinb. Sect. B Biol. Sci. *104*, 75–176.

[8] Brooks, T.M., Mittermeier, R.A., Fonseca, G.A.B. da, Gerlach, J., Hoffmann, M., Lamoreux, J.F., Mittermeier, C.G., Pilgrim, J.D., and Rodrigues, A.S.L. (2006). Global Biodiversity Conservation Priorities. Science *313*, 58–61.

[9] Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., and Kreft, H. (2005). Global Centers of Vascular Plant Diversity. Nova Acta Leopoldina NF 92 *342*, 61–83.

[10] Stropp, J., Ladle, R.J., M. Malhado, A.C., Hortal, J., Gaffuri, J., H. Temperley, W., Olav Skøien, J., and Mayaux, P. (2016). Mapping ignorance: 300 years of collecting flowering plants in Africa. Glob. Ecol. Biogeogr. *25*, 1085-1096.

[11] IUCN. The IUCN Red List of Threatened Species Version 2015-4, 2016. http://www.iucnredlist.org/#.

[12] Baksh-Comeau, Y.S., Maharaj, S.S., Adams, C.D., Harris, S.A., Filer, D.L., and Hawthorne, W.D. (2016). An annotated checklist of the vascular plants of Trinidad and Tobago with analysis of vegetation types and botanical "hotspots." Phytotaxa *250*, 1–431.

[13] Küper, W., Sommer, J.H., Lovett, J.C., Mutke, J., Linder, H.P., Beentje, H.J., Rompaey, R.S.A.R.V., Chatelain, C., Sosef, M., and Barthlott, W. (2004). Africa's Hotspots of Biodiversity Redefined. Ann. Mo. Bot. Gard. *91*, 525–535.

[14] Williams, P., Gibbons, D., Margules, C., Rebelo, A., Humphries, C., and Pressey, R. (1996). A Comparison of Richness Hotspots, Rarity Hotspots, and Complementary Areas for Conserving Diversity of British Birds. Conserv. Biol. *10*, 155–174.

[15] Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., and Gascon, C. (2011). Global Biodiversity Conservation: The Critical Role of Hotspots. In Biodiversity Hotspots, F.E. Zachos, and J.C. Habel, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 3–22.

[16] Linder, H.P. (2001). Plant diversity and endemism in sub-Saharan tropical Africa. J. Biogeogr. *28*, 169–182.

[17] Wieringa, J.J., and Poorter, L. (2004). Biodiversity hotspots in West Africa; patterns and causes. In Biodiversity of West African Forests: An Ecological Atlas of Woody Plant Species, L. Poorter, F. Bongers, F. N'. Kouamé, and W.D. Hawthorne, ed. (CABI), pp. 61-72.

[18] Lovett, J.C., Rudd, S., Taplin, J., and Frimodt-Møller, C. (2000). Patterns of plant diversity in Africa south of the Sahara and their implications for conservation management. Biodivers. Conserv. *9*, 37–46.

[19] White, F. (1993). The AETFAT Chorological Classification of Africa: History, Methods and Applications. Bull. Jard. Bot. Natl. Belg. Bull. Van Natl. Plantentuin Van Belg. *62*, 225.

[20] Lovett, J.C. (1998). Importance of the Eastern Arc Mountains for Vascular Plants. J. East Afr. Nat. Hist. *87*, 59–74.

[21] Thulin, M. (1993). Flora of Somalia, Volume 1 (Royal Botanic Gardens, Kew).
[22] Moore, J.J. (1962). The Braun-Blanquet System: A Reassessment. J. Ecol. *50*, 761–769.
[23] Hunter, M.L., and Hutchinson, A. (1994). The Virtues and Shortcomings of Parochialism: Conserving Species That Are Locally Rare, but Globally Common. Conserv. Biol. *8*, 1163–1165.
[24] Grenyer, R., Orme, C.D.L., Jackson, S.F., Thomas, G.H., Davies, R.G., Davies, T.J., Jones, K.E., Olson, V.A., Ridgely, R.S., Rasmussen, P.C., et al. (2006). Global distribution and conservation of rare and threatened vertebrates. Nature *444*, 93–96.
[25] Cincotta, R.P., Wisnewski, J., and Engelman, R. (2000). Human population in the biodiversity hotspots. Nature *404*, 990–992.
[26] O'Connor, C., Marvier, M., and Kareiva, P. (2003). Biological vs. social, economic and political priority-setting in conservation. Ecol. Lett. *6*, 706–711.
[27] Loreau, M. (2001). Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. Science *294*, 804–808.
[28] Vane-Wright, R.I., Humphries, C.J., and Williams, P.H. (1991). What to protect?—Systematics and the agony of choice. Biol. Conserv. *55*, 235–254.
[29] Linder, H.P., Lovett, J.C., Mutke, J.M., Barthlott, W., JÜRGENS, N., Rebelo, T., and Küper, W. (2005). A numerical re-evaluation of the sub-Saharan phytochoria of mainland Africa. Biol. Skr. *55*, 229–252.
[30] Küper, W., Sommer, J.H., Lovett, J.C., and Barthlott, W. (2006). Deficiency in African plant distribution data – missing pieces of the puzzle. Bot. J. Linn. Soc. *150*, 355–368.

Figure legends

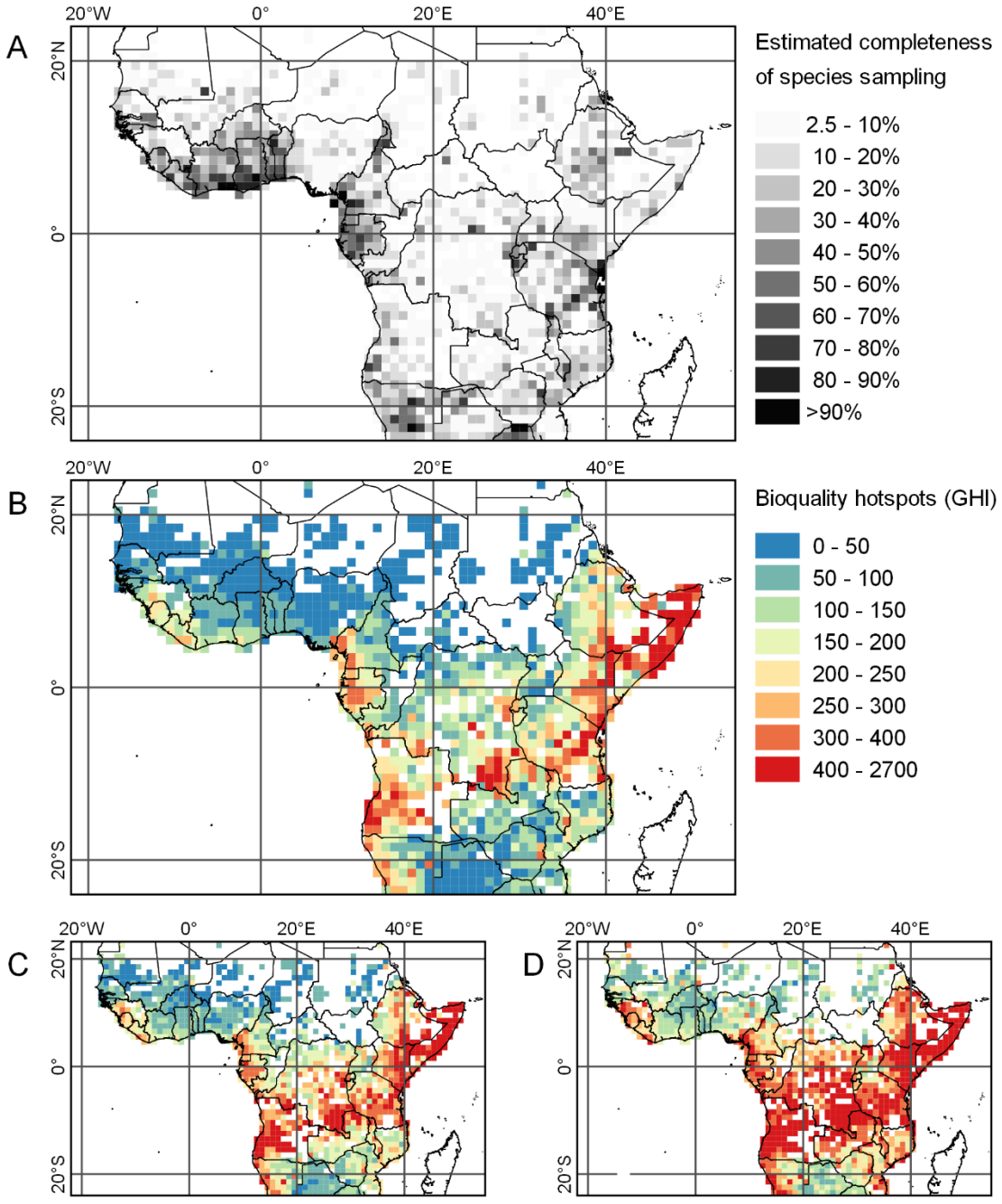


Figure 1. Bioquality hotspots in the tropical African flora. A: Ratio of species richness in our database to total species richness estimated using Barthlott et al. 2005 [9]. B: Bioquality mapped at one degree square resolution using minGHI, a reliable minimum estimate of GHI; minGHI is a conservative GHI estimate expected to be closer to the true GHI if collections are currently biased towards globally rare species. C: GHI values for bioquality (assumes no species sampling bias with respect to Star). D: maxGHI, maximum likely GHI assuming species sampling is currently biased towards the globally commonest species (probably the least likely scenario). Confidence intervals (minGHI to maxGHI) are larger where species sampling is poorer (compare panels A, B, C, D). 'True' GHI values, assuming perfect collection, would fall between minGHI and maxGHI estimates for each cell. See also Tables S1 and S2 Excel files.

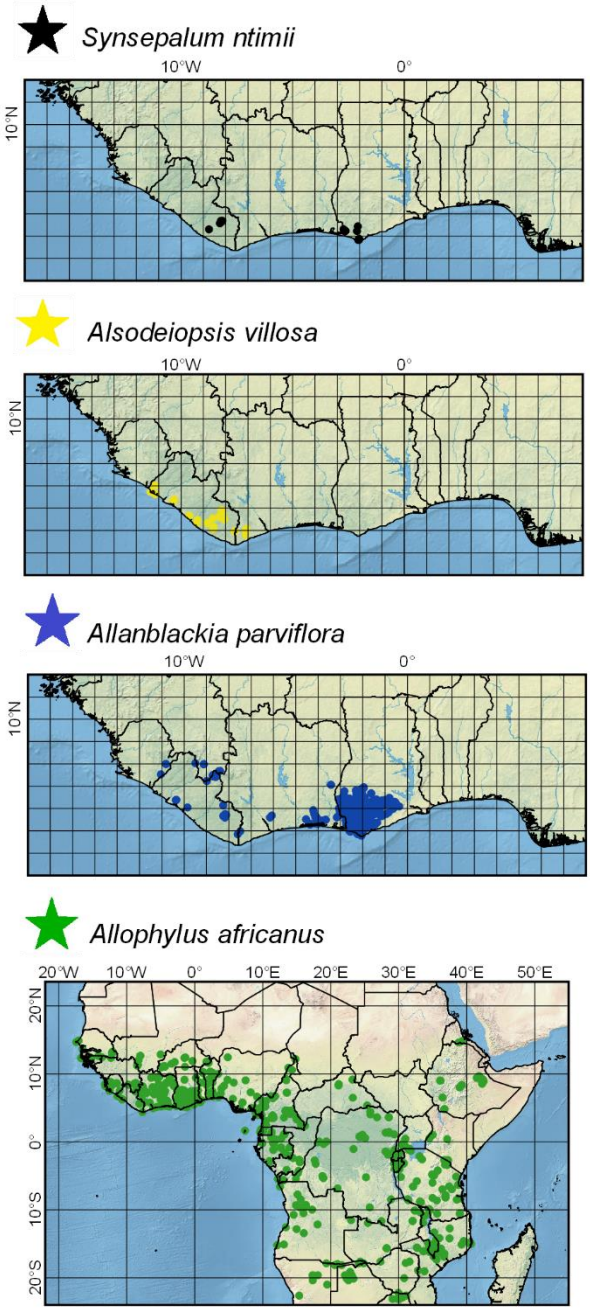


Figure 2. Example distribution patterns for a species of each Star. Black Star species occupy on average 2.7 degree squares globally. Gold Star species occupy 8, Blue Star species occupy 24, and Green Star species occupy 72 degree squares globally (or 100 x 100 km, whichever is the larger). Mapped distribution for *Allophylus africanus* includes distribution data for named formas and varieties. See also Table S1 Excel file.

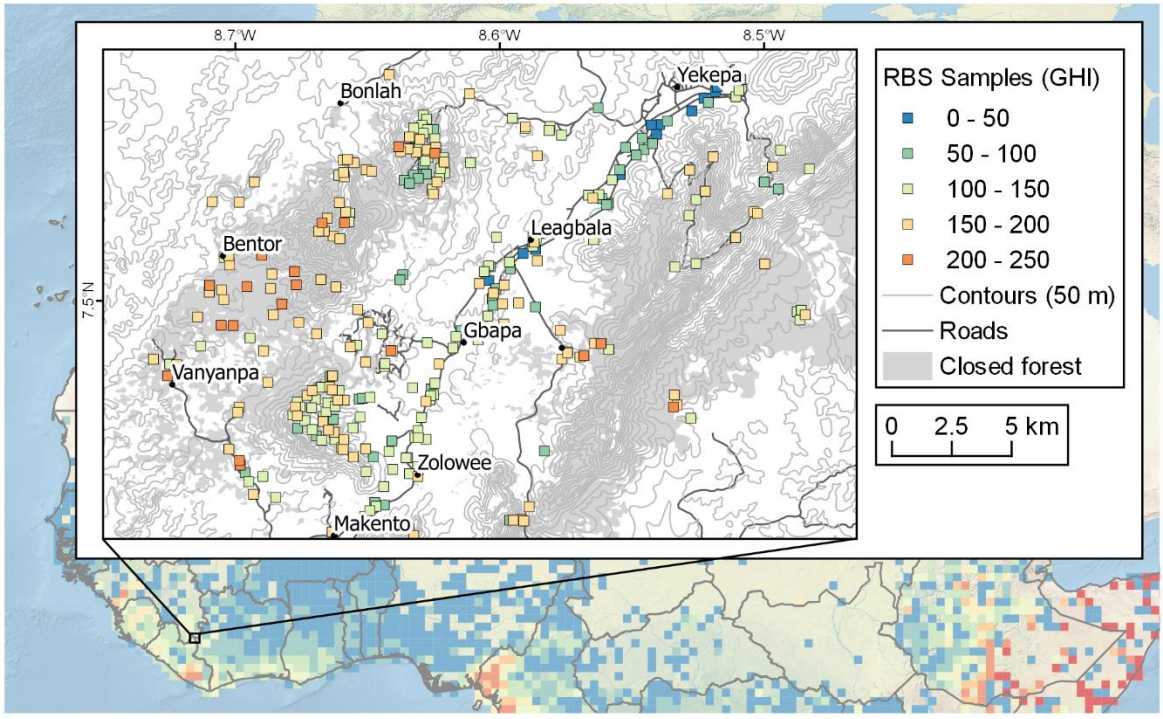


Figure 3. Bioquality at local scales. GHI calculated from 310 Rapid Botanic Survey (RBS) samples across northern Nimba County, Liberia. The hotter GHI (>200) scores equivalent to the minGHI estimate for the degree square as a whole were recorded in forest in this region, although not all the forest had such a high GHI. Background map shows minGHI for 0.5 x 0.5 degree squares. See also Tables S1 and S2 Excel files.

Table S1 (separate Excel file). Related to Figures 1, 2 and 3.
Star ratings for tropical African species.

Table S2 (separate Excel file). Related to Figures 1, 2 and 3.
Sampling levels and GHI scores for each cell in the analysis.

Supplemental Experimental Procedures

Taxonomic & geographic scope
We included any vascular plant collected on the mainland of Africa between the tropics of Cancer and Capricorn. We limited it to vascular plants because other plant groups are too poorly known to accurately assess their distribution and taxonomic status. We included native, naturalised, and introduced species in our database and Star rating, but did not include records of the most anthropogenic species (i.e. GX Stars, see Star Rating section) in bioquality calculations.

Species names & synonymy data
We assembled a database of tropical African plant species and their distributions using BRAHMS v7 [S1]. The initial species list was derived from the tropical African section of the African Plants Database (APD) [S2] and was transmitted to C.M. by Cyrille Chatelain in September 2014. Species names and synonyms were added to the database iteratively with distribution records and following new publications. We harmonised preferentially with Kew's World Checklist of Selected Plant Species (WCSP) [S3] for the Cyperaceae, Orchidaceae and Poaceae, and for Pteridophytes we followed the taxonomy presented in the Flora of Tropical East Africa (FTEA) via JStor Global Plants [S4] as far as possible. Record identifications were updated against this framework so that all records of a taxon were united under their accepted name. Species records without infraspecific names were taken to apply to the species as a whole: only records that were explicitly identified to infraspecific level were treated as such. In a few cases, the infraspecific taxon is implied by the geographical location, so in these cases the implied infraspecific taxon identity was applied.

Distribution data
We compiled distribution data from many different sources. The tropical African database (TRAFRICA) began with a West African database curated by W.H. These records came from survey work conducted in the region, and were enriched by European Community supported ECOSYN project (1996), which digitised many West African herbarium records. The Hall & Swaine database of Ghanaian records was added [S5]. C.C. sent C.M. records of species from Cote d'Ivoire, and country-level distribution data for the tropical African flora. Jan Wieringa transmitted the Nationaal Herbarium Nederland (NHN) BRAHMS database in its entirety to C.M., once in October 2014, once in September 2015 after NHN's digitisation efforts were complete, and a final update in June 2016. Distribution data for tropical African species were extracted from the database by C.M. Herbarium records from the NHN collections are available online [S6,7]. The BIOTA-BISAP plant distribution dataset was transmitted to C.M. in October 2014. This dataset is entirely gridded at a one degree square resolution (South African records were excluded).

Distribution data from the Global Biodiversity Information Facility (GBIF) [S8] were incorporated into the database. Data download was broken up geographically and taxonomically. Only vouchered records were downloaded, and only records with 'no known geographic issues' for records with coordinates. Other data sources included a mixture of observations and vouchered records. Downloads in Darwin Core Format (DwCA) were formatted for BRAHMS by C.M. These archived links include the contributors to the dataset for each search:

GBIF.org (3rd July 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.cw8ol3
GBIF.org (3rd July 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.wc8wzw
GBIF.org (21st July 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.jbojyk
GBIF.org (19th January 2016) GBIF Occurrence Download http://doi.org/10.15468/dl.wjnsjs
GBIF.org (17th March 2016) GBIF Occurrence Download http://doi.org/10.15468/dl.yuylfv
GBIF.org (17th March 2016) GBIF Occurrence Download http://doi.org/10.15468/dl.1l8wgx

Distribution records for African conifers were extracted from the BRAHMS database maintained by Aljos Farjon by C.M. with permission [S9], and for Mount Mulanje, Malawi, created by Alison Strugnell [S10]. A number of published checklists were included by formatting the pdf text to dbf files for import [S11–14].

Records for tropical African species from outside of tropical Africa were gridded at the one degree square resolution for the purpose of Star rating and were not submitted to the main database (c. 700,000 records).

Species were summarised uniquely by one degree square (Figure 1) for the analysis. Excluding straddling records (where location is defined by a polygon and that polygon straddles a sampling grid line), unlocalisable records, duplicate records (same species same degree square), records not identified at or below species level, and species with uncertain Stars or taxonomic status, resulted in 498,701 records for the one degree square summary.

The main database holds 3,013,061 distribution records.
Summary of TRAFRICA records by dataset:

Dataset	**Geographic coverage**	**Number of records in TRAFRICA**	**Provided by**
GBIF records	Tropical Africa, Global	1,659,027	Various: see text
WAfrica	West African	514,912	University of Oxford Dept of Plant Sciences
Hall & Swaine 1981	Ghana	19,132	M.D.Swaine *pers.comm*. to W.H. (curated by University of Oxford Dept of Plant Sciences)
NHN database	Global	493,729	NHN
WCSP country level	Global	80,314	WCSP
BIOTA-BISAP	Africa	73,759	Uni. Bonn
CJB Ivory Coast	Ivory Coast	64,054	CJB
CJB country level	Africa	62,236	CJB
Nimba survey data	Local: Nimba County	31869 8005	ArcelorMittal Liberia Euronimba Liberia Ltd (curated by University of Oxford Dept of Plant Sciences)
Checklists & smaller databases	Local: Various	6024	Various: see text & references

Readers should defer to the data providers for terms of reuse.

Cleaning geographic data
A number of routines were written in Microsoft Visual Foxpro 9 and Manifold GIS 8 to manage geolocation information. Place name dictionaries were imported into Foxpro and Manifold GIS from public domain gazetteers. Core spatial dictionary reference files were compiled in Manifold GIS version 8. Polygons for African administrative regions were downloaded from http://www.gadm.org/. These were processed by GIS, assigned unique geocodes and split into 6 layers depending on the administrative hierarchy. WDPA Protected Areas polygons were also processed, as defined by the World Conservation Union and UNEP-World Conservation Monitoring Centre's World Database on Protected Areas (2007, WCMC, Cambridge, UK, 2007). This dataset is available through the Global Land Cover Facility, www.landcover.org. Individual polygons in each layer were coded with the geocode of all polygons at all coarser scales in which they were entirely enclosed. African place names and point localities were downloaded from: http://www.geonames.org/. These were allocated unique geocodes and also assigned geocodes of all polygons in which they were enclosed at various levels.

Distribution database records were divided into those with specified coordinates, and those without. About 500,000 of the original distribution records did not have coordinates: The text locality information was compared against the core gazetteer dictionary files in a geocoding procedure to assign point-with-radius coordinates, or area coordinates where the specimen was not finely localised (e.g. only localised to a district, or country). Using these procedures we were able to assign area or point with radius coordinates to around 98% of them, although in many cases this was for larger administrative region only which do not fit into single degree square cells. For records with coordinates, the coordinates were reverse geocoded to assign all hierarchical place names to the record, and the text locality information was also geocoded to assign point or area coordinates. Where the two sources of information for a distribution record conflicted, most records were checked and corrected by hand where the problem was detectable, with the remainder being omitted from the analyses. All

analysis and mapping was conducted by asking whether the record's area bounding box (or point with radius) fitted entirely inside the sampling area of interest, rather than by assigning points to the middle of the box as is commonly done, so that the geographic resolution at which the record was originally collected could be respected.

We manually checked 2,033,783 of the records (67% of all records) for geographic accuracy. We achieved this by first making a file of unique locations (text locality-coordinate combinations). We checked: each unique gazetteer with more than 25 records; localities which are typically problematic (for example country centres); localities where stated coordinates conflicted with stated text locality information (as above); and many probable coordinate errors in passing. Also included in these 2 million checked records are records for which we trusted the collectors (e.g. our own samples located by GPS or the BIOTA dataset). This leaves a significant number of original records unchecked by us, but at least the locality records with the greatest impact on the Figure 1 result have been checked. For many localities with few botanical records and no supplied textual locality information, we, effectively, trusting the original collectors and data providers, having taken reasonable precautions to identify situations where mistakes would have been made. 18,294 records that were otherwise usable were lost because they straddled a sampling (grid cell) border.5684 records proved totally unlocalisable, and many others were reduced to country or top administration level areas (e.g. Western Region).

Our cleaned version of the original botanic records are available to the original data providers on request. Requests should be addressed to the corresponding author.

Star rating

Tropical African species were initially assigned a Star rating based on their global degree square occupancy, and the grid was defined with its origin at the meridian/equator intersect. Star rating followed the principle that Black Star species occupy on average 2.7 degree squares globally and carry a weight of 27; Gold Star species occupy on average 8 degree squares and carry a weight of 9; Blue Star species occupy on average 24 degree squares and carry a weight of 3; and Green Star species occupy on average 72 degree squares and carry a weight of 0. Species which were not believed to regenerate in natural vegetation in tropical Africa – widely cultivated species – were assigned the Star 'GX'. Because digitised distribution data are not complete, each species' Star was reviewed by the authors in light of the best available information from online floras and other botanic resources. Online resources consulted included the Plant List [S15], the JStor Global Plants Initiative [S16], the African Plants Database [S2,S17], WCSP [S3], IUCN Red List [S18], Tropicos [S19], and original publications, particularly for recently described species.

Digitised distribution records are not yet sufficiently comprehensive to accurately reflect all plant species' true distributions globally. It is known that certain groups of species are under recorded in herbaria, because of their geography or ecology. Fortunately, additional information about species' plausible distributions is available in formats other than gridded distribution data, such as those described above. A categorical system is better able to take advantage of this information than a continuous system because it is possible for a knowledgeable person to estimate which broad category of global rarity a species is likely to belong to, but not to assign a precise number of degree squares of occupancy. Although this introduces a degree of subjectivity to the system, the results are more accurate and better reflect the true breadth of knowledge regarding species' distributions than a strict reliance on digitised records would, and allows biases in a species' recorded collections to be compensated for. Occasionally, species are too poorly known to allow a Star to be assigned, and they are designated with '?'.

Star ratings are updated following available information and new species are regularly Star rated. The latest Star ratings for species worldwide can be found on our Star Server [S20]. A detailed discussion of Star Rating can be found in the RBS manual [S21]. The Stars as derived for this publication and used in this analysis are provided as a download (Table S1).

Calculating GHI and confidence intervals

GHI is calculated using the following formula, where N_{BK}, N_{GD}, N_{BU} and N_{GN} are the number of Black, Gold, Blue and Green Star species, and W_{BK}, W_{GD} and W_{BU} are the respective weights.

$$GHI = 100 * (N_{BK} \times W_{BK} + N_{GD} \times W_{GD} + N_{BU} \times W_{BU})/(N_{BK} + N_{GD} + N_{BU} + N_{GN})$$

GX species are invisible in this GHI calculation. GX species are species whose distributions are largely if not entirely anthropogenic in tropical Africa (species like cassava or *Ixora coccinea)*, and incidentally are not regularly collected by botanists. These records were excluded from GHI calculations so that the GHI of a cell would not be unfairly brought down by the presence of a city, farm, or botanic garden (for example) in the cell.

Although these species are plants, there are treated here like a statue would be – invisible – because we are concerned with documenting bioquality in natural vegetation, and not the bioquality of gardens (which is an interesting but separate matter). We have run the same analyses including GX species, and find there is almost no perceptible difference in the overall patterns, but cells containing well documented botanic gardens (e.g. Limbe botanic garden, Cameroon) have a slightly lowered GHI when GX species are included).

Confidence intervals were generated using a resampling procedure. Ideally, GHI is based on all or at least >90% of the flora of a place. For local samples, this is achievable. At the one degree resolution, we know that we have <90% of the species recorded for most cells, and we do not know whether we have a biased (with respect to Star) sample of the flora. The bootstrapping approach described here does not test for these biases. Rather, it was used to create confidence intervals around the GHI scores, such that if there were a bias in either direction for a cell, the true value would still be within the confidence interval.

We estimated that 7 degree squares in tropical Africa had >90% of their species recorded. For each such 'sub-complete' cell, we calculated the 'true' GHI (set GHI), using all the species present. Then, we drew 2.5% of those species and recalculated the GHI of the small sample. We repeated this 10,000 times, and from this batch of 10,000 sample GHIs we calculated the 1st and 99th percentiles of the GHIs. We plotted these min and max points against the true set.ghi; we did this for all 7 cells; and then we modelled linearly the min and max values across the range of the GHI. This allows setGHI to be estimated for any sampleGHI; each cell has a different balance of Stars and number of species, and this influences the maxGHI-minGHI ranges; the linear models are a best fit to this variation. We repeated this, drawing 5%, 10%, 20%, 30%, 40%, 50%, 75%, 90% of the species; as a greater proportion of species are drawn, the range of values that sampled GHIs can take gets much more constrained by the true value of the cell. Having established these relationships, we treated each cell as a sample GHI with a particular sampling proportion, and asked what values the true set GHI could take, given the sample GHI and the sampling proportion (sampling proportion defined as in Figure 1A). For example, a cell with 50 species recorded from it where 2000 species might be expected has an estimated sampling intensity of 2.5%. If the GHI calculated from those 50 species is 500 (the sample GHI), then the 'true' GHI of the cell is likely to fall between 276 and 920. 276 is the minimum likely true GHI for the cell, given the sample GHI of the 50 species present. MinGHI was not mapped for any cell where <2.5% of the species were estimated to have been recorded.

Total species richness for each one degree cell was estimated using [S22]. The map was digitised, and each degree square was assigned to the maximum Diversity Zone which it overlapped, corresponding to an estimated species richness at approximately this scale. This map was created using a much smaller dataset, and reports species richness estimates at 10,000 km^2 resolution rather than 1 degree square, but it serves as a reasonable approximation for this purpose. We estimated sampling completeness by comparing our recorded richness against these estimates and presented the results as a percentage. To estimate total richness for areas of different sizes (for example in the quarter degree squares, some of which are visible in Figure 3), we did the following: We assigned each degree square to a simplified version of the African phytochoria [S23]. We took those estimates of total species richness for each phytochorion. We averaged species richness per one degree cell for each phytochorion. A dummy species list was created of equal length to the total richness values for each phytochorion, this pool was resampled by drawing the average estimated number of species per one degree square per phytochorion; the number of unique dummy species in each 2, 3, 4 and 5 degree square area was calculated and a species area curve fitted to the values, to create theoretical species area curves for each phytochorion. This method is crude, as it does not account for any variables affecting species richness beyond the size of area and the phytochorion, but produces plausible richness estimates for the purpose of constructing confidence intervals around GHI at scales other than one degree square. The phytochorion shape file is available at
http://www.arcgis.com/home/item.html?id=9d465d61559a414fb866a50e0f09235c

GHI results at the one degree square are provided as a download (Table S2).

Mapping
Mapping was conducted using QGIS (http://www.qgis.org/en/site/). Geocoding and reverse geocoding analyses were conducted in part using Manifold (http://www.manifold.net/index.shtml).

Fieldwork
The survey data shown in Figure 3 were collected using the RBS method [S21]. RBS has been refined over 25 years to fill the distribution data gap which exists between herbarium records and formal (tree-) plot records. Survey took place between 2010 and 2012 as part of ArcelorMittal Liberia's (AML) Environmental and Social

Impact Assessment (ESIA) [S24] and an ESIA of the rail corridor for Euronimba Liberia Ltd. 31,775 records were made of plants across 310 RBS samples. Species were identified by W.H. and C.M, Pierre Poilecot and Ouo-ouo Haba, Patrick Ekpe, Carel Jongkind, James Kpadehyea, David Bilivogui, Steven Heathcote, Wing-Yunn Crawley and Daniel Dorbor. Specimens were deposited at FHO (Oxford) and with AML pending the resurrection of Liberia's National Herbarium.

Supplemental References

[S1] Department of Plant Sciences, University of Oxford (2013). Botanical Research and Herbarium Management System: Documentation - BRAHMS Online. http://herbaria.plants.ox.ac.uk/bol/brahms/Documentation (accessed 2014-2016).

[S2] Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria. African Plants Database (version 3.4.0) 2016. http://www.ville-ge.ch/musinfo/bd/cjb/africa/ (accessed 2014-2016).

[S3] Govaerts R et al. World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. 2016. http://apps.kew.org/wcsp (accessed 2014-2016).

[S4] Royal Botanic Gardens, Kew. The Flora of East Tropical Africa (FTEA) 1948-2012. https://plants.jstor.org/collection/FTEA (accessed 2014- 2016).

[S5] Hall, J.B., and Swaine, M.D. (1981). Distribution and ecology of vascular plants in a tropical rain forest (The Hague - Boston - London: Dr W Junk Publishers).

[S6] Naturalis Biodiversity Center. Bioportal 2016. http://bioportal.naturalis.nl/.

[S7] NHN: BRAHMS Online 2016. http://herbarium.naturalis.nl/

[S8] GBIF: Free and Open Access to Biodiversity Data 2016. http://www.gbif.org (accessed 2015-2016).

[S9] Farjon A. Conifers of the World: Resources for Conifer Research 2016. http://herbaria.plants.ox.ac.uk/bol/conifers (accessed 2016).

[S10] Strugnell, A.M. (2006). A Checklist of the Spermatophytes of Mount Mulanje, Malawi. Scripta Botanica Belgica 34 (National Botanic Garden of Belgium, Meise, Belgium).

[S11] Ouédraogo, O., Schmidt, M., Thiombiano, A., Hahn, K., Guinko, S., and Zizka, G. (2011). Magnoliophyta, Arly National Park, Tapoa, Burkina Faso. Check List *7*, 85–100.

[S12] Mbayngone, E., Schmidt, M., Hahn-Hadjali, K., Thiombiano, A., and Guinko, S. (2008). Magnoliophyta of the partial faunal reserve of Pama, Burkina Faso. Check List *4*, 251–266.

[S13] Assédé, E.P.S., Adomou, A.C., and Sinsin, B. Magnoliophyta, Biosphere Reserve of Pendjari, Atacora Province, Benin. Check List *8*, 642–666.

[S14] Ismail, I.M., and Elawad, A.A. (2015). Checklist of plants of Rashad and Alabassia localities (eastern Nuba Mountains), South Kordofan Sudan. Check List *11*, 1805.

[S15] The Plant List Version 1.1 2016. http://www.theplantlist.org/. (accessed 2016)

[S16] Global Plants on JSTOR 2016. http://plants.jstor.org/.

[S17] Lebrun, J.P., and Stork, A.L. (1991). Tropical African Flowering Plants: Ecology and Distribution (Geneva: Conservatoire et Jardin botaniques de la Ville de Genève, Geneva, Switzerland).

[S18] IUCN. The IUCN Red List of Threatened Species Version 2015-4 2016. http://www.iucnredlist.org/# (accessed May 2016).

[S19] Missouri Botanic Garden. Tropicos 2016. http://www.tropicos.org/. (accessed 2014-2016).

[S20] Department of Plant Sciences, University of Oxford. The Oxford Plant Observatory TOPO 2016. http://herbaria.plants.ox.ac.uk/bol/topo/

[S21] Hawthorne WD, Marshall CAM. A Manual for Rapid Botanic Survey (RBS) and measurement of vegetation bioquality 2016. http://herbaria.plants.ox.ac.uk/bol/oxford/Survey.

[S22] Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., and Kreft, H. (2005). Global Centers of Vascular Plant Diversity. Nova Acta Leopoldina NF 92 *342*, 61–83.

[S23] White, F. (1983). Vegetation of Africa - a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa; Natural Resources Research Report XX.

[S24] URS. Volume 4, Part 1.1: Forest Botanical Baseline & Impact Assessment 2013. http://liberia.arcelormittal.com/corporate-responsibility/publications-and-reports.aspx (accessed May 27, 2016).